DOCUMENT RESUME

ED 118 258 95 PS 008 353

AUTHOR Webster, Murray, Jr.
TITLE Expectation Effects on Performance Evaluations.
SPONS AGENCY Office of Education (DHEW), Washington, D.C.
PUB DATE Aug 75
GRANT OEG-3-71-0122
NOTE 24p.; Paper presented at the Annual Meeting of the American Psychological Association (83rd, Chicago, Illinois, August 30 - September 3, 1975)

EDRS PRICE MF-$0.83 HC-$1.67 Plus Postage
DESCRIPTORS Ability Identification; *Bias; *Elementary Education; *Expectation; Grade 3; Situational Tests; *Student Ability; *Student Evaluation; Task Performance
IDENTIFIERS *Expectation States Theory

ABSTRACT

This study examined the effects of varying degrees of situational definition on expectation states of third graders. The experiment was conducted in the spring and replicated in the fall. A total of 389 children in 12 third grade classrooms participated. The experimental conditions reflected three degrees of situational definition. In the assignment condition, subjects were told that another child had high or low ability at a certain task and that this ability was relevant to the task at hand; then expectations for the task at hand were assigned to the other child. In the relevance condition, subjects were told that the other child had high or low ability at another task and that this ability was relevant to the task at hand, but no assignment of expectations to the other child was made. In the activation condition, subjects were told only the other's ability at another task. Each child was then asked to rate 10 sentences in which some words had been supplied by the other (fictitious) student. The results indicated that in both experiments, subjects in the activation and assignment conditions reflected about equal expectation effects in their ratings, while those in the relevance condition showed a smaller expectation effect. It was provisionally concluded that the data support the equality assertion of the expectation states theory. (JMB)

****
* Documents acquired by ERIC include many informal unpublished *
* materials not available from other sources. ERIC makes every effort *
* to obtain the best copy available. Nevertheless, items of marginal *
* reproducibility are often encountered and this affects the quality *
* of the microfiche and hardcopy reproductions ERIC makes available *
* via the ERIC Document Reproduction Service (EDRS). EDRS is not *
* responsible for the quality of the original document. Reproductions *
* supplied by EDRS are the best that can be made from the original. *
****

ED118258

U.S. DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION
THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

EXPECTATION EFFECTS ON PERFORMANCE EVALUATIONS

Murray Webster, Jr.

University of South Carolina

July, 1975

PS 008353

Abstract

Situations in which expectations affect performance evaluations of actors are described and analyzed in terms of expectation states theory. Results of three experiments with grade school children (n = 289) are presented testing derivations of that theory, and some implications of the results for theory building and for practical applications are discussed.

EXPECTATION EFFECTS ON PERFORMANCE EVALUATIONS [1]

Evaluations of performance, and by extension, of the actors making those performances, are important in a wide variety of social situations. Of particular interest to social psychologists are cases in which evaluations are influenced by factors other than objective criteria. A variety of processes--"halo effect" (Symonds, 1925), "cognitive set" (Kelman, 1961), group influence (Asch, 1956), and others--have been proposed to account for discrepancies between objectively recorded performances and subjective evaluations of performance.

Distortion of evaluations in the direction of previous beliefs about abilities of the performing actors has been widely documented (see, for example, Bales, 1970; Rist, 1970; Kelman, 1961). If one actor is thought to be good at a task, his performances are more likely to be highly evaluated than equal performances by an actor thought to have low task competence. One way to account for this type of cognitive distortion has been proposed by Berger, et al. (1972). They argue that problem solving interaction leads to formation of more or less enduring ability conceptions, called expectation states, for the actors involved. Once expectation states are formed and attached to actors, they are predicted to affect most important features of behavior: the higher the expectations held for a given actor, the more likely he is to be given and to accept chances to perform, to receive agreement and positive evaluations for any of his performances, and to exert influence over other group members.

Although this theory can account for many of the evaluation distortions previously mentioned, and in fact is claimed to apply to an even wider range of phenomena, direct tests and applications of the theory

are still few. At present all direct tests have been conducted in laboratory settings, and every one of these tests has taken as its dependent variable rejection of influence in case of disagreement, only one of the many behaviors predicted to vary with expectation states. This limitation makes it hard to assess the many untested predictions of the theory, or its practical usefulness.

Knowledge of another's expectation state, or of a way to change it, would have enormous value in day-to-day living. Accordingly, the experiments reported here use a naturalistic setting and test how previous expectations affect future evaluations of performance. They also extend our previous work on the determinants and consequences of school children's expectations for their own performance, and how to change them (see Entwisle & Webster, 1972, 1974, for summary).

Much of the interaction in classrooms is concerned directly with the performances, evaluations, and expectations for future performance which are the major variables of expectation states theory. Teachers ask questions in class, give tests, and make homework assignments; students respond and their performance attempts are evaluated by the teacher and by other students. Moreover, evaluations of schoolchildren and their performances cannot be wholly based on objective criteria. For one thing, objective evaluation standards are often lacking or ambiguous. What, for instance, differentiates a good poem or a good painting from a bad one? Second, correct answers are hard to recognize immediately for difficult or complex tasks. Grading algebra problems, or evaluating a single line of a computer program is difficult for this reason. In elementary school, peer evaluations are very important to students, but they often come from peers who are not capable of recognizing good or

bad performances. Third, teachers often must evaluate students whom they have not had sufficient opportunity, or perhaps any opportunity, to test adequately. Marks in any subject are often based on inadequately sampled classroom behavior. Peer evaluations likewise are frequently based upon inadequate or biased test information. Fourth, often teachers are required to evaluate kinds of performance that neither they nor students can define. First grade teachers, for example, often give marks in "Language" or in "Spelling" when neither they nor their first graders can say exactly what "Language" or "Spelling" involves at this level. Finally, expectations for some students are so firmly fixed on prior "evidence" that performance not in accord with expectations is disbelieved (a phenomenon documented recently for first-grade children by Seaver, 1973, and earlier by Whyte, 1943, in the famous bowling incident of Street Corner Society). Each of the above circumstances can operate to attenuate evaluation solely by objective standards. To the extent that objective criteria are not used, expectations held for individuals making the performances will influence evaluations.

If objective evaluative criteria fail to specify evaluations exactly, then three possible situations, differing in degree of social definition provided by expectation information, are possible: (1) In a maximally defined situation, performance expectations for the specific task at hand have been previously assigned to actors. For example, an actor may be known to posses high ability at task A, which is just the skill needed for successful task completion. Then it is a simple step to decide that any specific performance of the actor is probably a good one and should be highly evaluated. One expects a Metropolitan Opera singer to do well on a given aria even though one has never heard him sing it before. (2) In a less completely defined situation, ability

is required at task A, but no expectations for performance of the actor at task A are known. However expectations are held for his performance at task B, and task B is known to be relevant to task A. The theory predicts that expectations will then be formed for task A which are the same as those already held for task B. If a person excels in snow skiing, we expect he will be good at water skiing. (3) In a minimally defined situation, ability at task A is required, ability at task B is known, but nothing is known about whether tasks A and B are related. In this case, so long as tasks are not explicitly dissociated (as they would be, for instance, by telling an actor that the skills needed are completely different), the prediction is that individuals will conclude that the tasks are probably related. A research assistant who is good at computer programming will be expected to be good at planning experiments if nothing else is known about him. This process yields the same outcome as types (1) and (2) situations: the tasks are assumed to be relevant, expectations are assigned to actors, and these expectations affect the likelihood that actors will receive positive evaluations of performances.

One interesting consequence is that all three situations are predicted to yield identical outcomes for both expectation states and subsequent behaviors of individuals. That is, so long as certain conditions are met, it does not matter whether the process is entered at stage (3) where only ability at task B is known, at stage (2) where in addition task B is known to be relevant to task A, or a stage (1) where ability at task A is also known. This consequence may be empirically accurate but the only data available for direct test (Berger et al., 1972) do not provide particularly strong confirmation. Three experimental studies of

the same theory are provided here for social situations comparable to those of Berger et al. The concrete settings, the experimental design, the subject population, and the measure of expectation states differ from those used previously, however.

Experimental conditions were designed to reflect the varying degrees of situational definition described above. Condition 1, which we call assignment, was maximally defined: subjects were told the ability of the other at certain tasks, this ability was made relevant to the particular task, and finally, expectations for the particular task were assigned to the other person. Condition 2, relevance, was less completely defined: subjects were told other's ability at another task and the ability was said to be relevant to the particular task, but no assignment of expectations to other was made. Condition 3, activation, was minimally defined: subjects were told only other's ability at another task, but no mention was made of possible relevance between tasks.

Experiments were conducted in the spring and replicated in the fall with different children in third grade classrooms of two suburban schools. The investigator addressed the entire class, and told the children he or she was interested in finding out how well they could tell good words from bad words in sentences--the operational measure of "unit evaluation of a performance," which is predicted to vary directly with expectations held for other. (Good words were described as those which fit in well with the rest of the sentence, and which are exciting.) Then we handed each child a sheet with 10 sentences (Chart 1). Each

---

Chart 1 here

---

sentence contained some underlined words ("performances") supposedly supplied by a (fictitious) other student. The fictitious students were described as having either high ability at schoolwork (Task B), or low ability. Instructions for the three conditions varied as follows.

> [all] We have some words given by boys and girls with very high ability at most school subjects, and some by boys and girls with low ability. Students with high ability do better at reading, spelling, and arithmetic than students with low ability.
>
> [relevance and assignment only] Students with high ability at schoolwork also usually give better words than students with low ability.
>
> [assignment only] Today we have some words from students who have given good words in the past, and some from students who have given poor words in the past.
>
> [all] However we have not yet graded the words you are about to see.

As mentioned, the set of experiments was run twice. Two classrooms of subjects were used for each condition on both occasions, making six classrooms per experiment, or twelve classrooms all told (the average of persons per classroom = 24.1). The sentences shown in Chart 1 were used in every classroom. In the first classroom of each condition, every odd numbered sentence was described as coming from a good student. In the second classroom, descriptions of the fictitious authors were reversed: the even numbered sentences were supposedly given by good

students. The high or low expectations induced for the author of each sentence thus contribute a treatment variable which we predict will have the major effect on performance evaluations given to the words.

Subjects graded each sentence by checking a box from the following:

very good good fair poor very poor.

Data were tallied by arbitrarily assigning "very good" a score of 1, and "very poor" a score of 5. Tables 1 and 2 present mean scores given to sentences, according to expectation treatment.

Tables 1 and 2 here

Table 3 presents results of the Wilcoxon test for differences between evaluations of sentences within each experimental condition. Every condition except relevance in Experiment 1 shows a satisfactory level of significance for the predicted expectation effect.

Table 3 here

Finally, we may examine these data for evidence regarding one of the more refined theoretical assertions. Berger et al. (1972) have argued that all three of the information conditions should lead to equal expectations. That is, whatever the amount of information individuals are given, they are predicted to form equivalent expectation states. In our experiments, this "equality assertion" leads to a prediction that expectation effects, or difference scores, will be equal across all three conditions. The most reasonable alternative to equality

of effect would be a prediction that expectation states will form as a direct function of the degree of relevant information provided individuals. If this happened, then our conditions should be ordered by magnitude of expectation effect as follows: activation ≈ relevance < assignment.

It is evident from overall mean values in Table I and II that the ordering did not obtain. In both experiments, activation and assignment produced about equal effects, and relevance produced less effect. However we may still attempt to assess the "equality assertion," by performing statistical tests upon pairs of conditions within each experiment. Table 4 presents results of Mann-Whitney U tests of difference between

---

Table 4 here

---

relevance and both other conditions.[2] None of these differences reaches a satisfactory level for concluding that the conditions do indeed differ in magnitude of effect.[3]

Discussion

These results show that, in three different information conditions expectations held for an individual affected evaluations of his performances. Similar effects have been remarked upon in other settings, but our work apparently constitutes the first direct test of such a prediction. The effect becomes more significant when placed in the context of a theory, and tested in a context where most other systematic sources of variance are excluded.

Results also show the considerable impact of expectations on peer

evaluations in the classroom. Because peer evaluations are important factors in development of self-evaluation, it is useful to confirm that the general relation predicted between expectations and evaluations holds in this specific case. The general laughter which greets perfectly good answers from some students, as well as the awe accorded to others, seem interpretable in terms of the theory.

Expectation states theory asserts that individuals will make the cognitive connections necessary to structure the incomplete relevance and activation conditions. As a result, these two conditions should produce expectations identical to those given to subjects in the assignment condition. Quite clearly in both experiments, the magnitude of expectation effect (difference scores) for activation and assignment are approximately the same. Relevance shows a smaller expectation effect in both experiments, though not enough of a difference to sustain a claim of any ordering--or even to be confident that any reliable difference exists. Provisionally, then, we conclude that the data support the equality assertion of the theory.

Our reservation, of course, comes from the fact that relevance produced a smaller expectation effect in both experiments. Tests performed for Table 4 do not show that this difference is significant, so it would not appear wise to spend too much time explaining it. Yet the number of subjects (289) and of judgments (10 by each) is substantial, and perhaps some other analysis of the data could make the difference seem more important. In fact, disregarding the expectation treatment, all sentences were judged more favorably in relevance than in the other conditions. The mean scores for all sentences in both

expectation treatments were: assignment, 2.72; relevance, 2.43; activation, 2.65. Rather than being theoretically significant, to us these data suggest that something about experimental technique or the setting may be responsible for the shift. Several possibilities exist.

What seems most likely is that something about the instructions for relevance elicited sympathetic or other feelings in subjects. In both experiments, sentence ratings for the low expectation treatment were higher in relevance than in either of the other conditions. Perhaps subjects felt that sentence authors in the low expectation treatment should be encouraged by giving them higher evaluations. Conversely, perhaps something about the relevance instructions elicited a desire to reject expectations as the basis for evaluating words. In this case, subjects may have concluded "Well, perhaps smart kids usually give better words, but that doesn't mean it is true in this case." Either response--attempt to motivate in the low treatment, or rejecting expectations as the basis for evaluations--would decrease the effectiveness of the relevance condition. Furthermore, our impression is that both responses are normative in grade school culture. Children are often told to "give the benefit of the doubt," when judging, and even that a negative evaluation is impolite. They are also told to judge each case on its merits, not on what is already believed about people. Just why relevance and not the other conditions should arouse either of these emotions is not clear. In assignment, perhaps it was impossible to ignore the expectations and their significance to the evaluating task. In activation, perhaps it was the near lack of useful information which directed attention towards expectation formation, and away from norms blocking their use.

A second possibility, already mentioned, is that expectation effects in relevance varied from the other conditions as a function of different experimenters or different subjects. This fluctuation is the sort statistical tests are designed to assess, and by that criterion (Table 4) the effects were not significant.

Finally, although by design the study was an experiment, the setting was naturalistic. Certainly we could not exercise control over all extraneous influences affecting the dependent variable. In other words, if the experiment were replicated, perhaps no differential effect in relevance would be observed.[4]

Besides the findings from these experiments, developing this new measure of expectation states (evaluations of others' performances) should increase the flexibility accorded researchers. The measure is easy to implement and has straightforward relations with many cases of practical interest where the established measure, disagreement resolution, is not practical or even possible to use. For example, in committees, conferences, work groups, there often are norms which discourage the appearance or the recognition of disagreements. Also, the evaluations measure may be used in cases where the individual is not in direct interaction with others for whom differential expectations are held; for example, where teachers evaluate their pupils, or where a supervisor evaluates employees, as in Caudill's (1958) hospital.

Finally, we wish to note that, although our results show a tendency to base evaluations partially on expectations rather than on objective criteria, the tendency is not always harmful. To say that evaluations "should be" independent of expectation states seems as

useless as saying that people "should be" unconcerned with status in their interactions. In many situations individuals must reach evaluative conclusions when they simply do not have access to complete information from objective sources. In these situations, expectations held for actors--which result in most cases from observing previous performances of these actors--may well constitute the most adequate available basis for structuring an underdefined situation. What is rightly objected to is over-reliance on expectations in situations where objective criteria for evaluation are readily available.

Footnotes

[1]Experiments are supported by Office of Education Grant no. 3-71-0122, awarded to Doris R. Entwisle, Principal Investigator. These experiments are part of a larger research program, and were designed and conducted in collaboration with Dr. Entwisle. Her major contributions to this phase of the project are hereby acknowledged. In addition, Dr. Evart Cornell and Mr. George Fanshaw, Princpals, and the staff and students of Loch Raven and Stoneleigh Elementary Schools generously provided their time and help.

[2]The relevance condition of Experiment I does differ significantly from both activation and assignment of Experiment II by this test. However such a difference could well have been produced by the three month interval between experiments, by the change in experimenters between relevance I (Entwisle) and Experiment II (Webster), or by some unknown difference between children in the different classrooms.

[3]Since the "equality assertion" is equivalent to a null hypothesis in this experiment, we accept it by this test. We do not, of course, wish to claim that the statistic supports the null hypothesis; only that we cannot reject it nor can we offer a more satisfactory prediction at this time. The fact that Berger et al.(1972) also report no difference in 3 comparable conditions of a different experimental setting with a different subject population (Air Force personnel) lends confidence to our interpretation.

[4]These sentences differ, both in their average scores regardless of expectation manipulation, and in their susceptibility to the expectation manipulation. Such variation, of course, is why we use ten sentences rather than only one for each experimental condition. It most likely indicates that individuals respond differently to the content of sentences, not that the individuals differ importantly in their baseline propensity to distribute positive evaluations. We would be more concerned by the latter possibility if students were assigned to classes on some systematic basis (such as ability tracking) which might be related to the dependent variable--which they were not.

Chart I

Sentences Used for the Evaluation Task

1. There once was a very tall prince.
2. There once was a handsome movie star.
3. In order to fool the ladies, the man dressed up as a piece of furniture.
4. In order to fool the robbers, the princess dressed up as a big black bear.
5. The ocean was full of sharks and whales.
6. The parking lot was full of people with dogs.
7. When the Indians found the cowboy, they sat down and said hi!
8. When the teacher found the book, she opened it and began reading.
9. This was a good thing to do on Sunday.
10. This was more than he wanted.

Table I

Experiment I: Mean Scores of Sentences

by Expectations for Author

	Assignment			Relevance			Activation		
(n's)	(25, 30)			(17, 32)			(28, 28)		
Sentence	High	Low	Difference*	High	Low	Difference	High	Low	Difference
1	2.20	2.70	+ .50	2.00	2.39	+ .39	2.00	2.50	+ .50
2	1.42	2.76	+1.34	1.81	2.31	+ .50	1.68	2.04	+ .36
3	3.25	3.80	+ .55	2.76	2.84	+ .08	3.18	3.82	+ .64
4	2.46	2.64	+ .18	2.13	2.56	+ .43	2.68	2.86	+ .18
5	1.83	2.68	+ .85	1.88	1.94	+ .06	1.43	2.64	+1.21
6	2.96	3.79	+ .83	2.84	2.82	- .02	2.46	2.78	+ .32
7	4.27	3.64	- .63	2.12	3.65	+1.53	3.29	4.07	+ .78
8	1.92	2.77	+ .85	2.00	1.82	- .18	2.04	2.44	+ .40
9	2.33	2.48	+ .15	2.47	2.41	- .06	2.37	3.00	+ .53
10	1.87	2.67	+ .80	2.41	2.35	- .06	2.14	2.86	+ .72
Overall Mean	2.45	3.00	+ .55	2.24	2.51	+ .27	2.32	2.89	+ .57

*positive difference indicates predicted direction

Table 2

Experiment II: Mean Scores of Sentences

by Expectations for Author

	Assignment			Relevance			Activation		
(n's)	(27, 28)			(25, 11)			(19, 19)		
Sentence	High	Low	Difference	High	Low	Difference	High	Low	Difference
1	2.04	2.79	+ .75	2.20	2.45	+ .25	2.05	2.26	+ .21
2	1.71	2.44	+ .73	1.64	2.52	+ .88	2.21	2.90	+ .69
3	2.52	3.79	+1.27	2.54	4.18	+1.64	2.68	4.47	+1.79
4	1.89	2.70	+ .81	1.91	2.44	+ .53	2.26	2.79	+ .53
5	1.38	2.57	+1.19	1.96	2.91	+ .95	2.53	3.24	+ .71
6	2.39	3.31	+ .92	2.82	3.32	+ .50	3.06	3.79	+ .73
7	3.11	4.39	+1.28	2.88	3.64	+ .76	3.28	3.89	+ .61
8	1.75	2.41	+ .66	1.55	2.12	+ .57	1.74	2.37	+ .63
9	2.18	3.21	+1.03	2.60	2.00	- .60	2.05	3.26	+1.21
10.	2.25	2.56	+.31	1.36	2.68	+1.32	2.11	3.47	+1.36
Overall Mean	2.12	3.02	+ .90	2.33	2.90	+ .57	2.40	3.24	+ .84

Table 3

Results of Wilcoxon Test Within Conditions

	Assignment		Relevance		Activation	
	T	p*	T	p	T	p
Experiment I	5	.01	13	.07**	0	<.01
Experiment II	0	<.01	5	.01	0	<.01

* one-tailed

** Z transformation=1.48

Table 4

Mann-Whitney U Tests of Difference of Conditions

	Experiment I		Experiment II	
Conditions:	relevance, assignment	relevance, activation	relevance, assignment	relevance, activation
U	30	25	42.5	48
Z-Transformation	1.51	1.89	.57	.15
p*	.13	.06	.57	.88

* two-tailed

References

Asch, S. E.

1956 Studies of independence and conformity: a minority of one against a unanimous majority. Psychological Monograph, 70, no. 9.

Bales, R. F.

1970 Personality and Interpersonal Behavior. New York: Holt, Rinehart & Winston.

Berger, J., B. P. Cohen, & M. Zelditch.

1972 Status characteristics and social interaction. American Sociological Review, 37 (June): 241-255.

Caudill, W. K.

1958 The psychiatric hospital as a small society. Cambridge: Harvard University Press.

Entwisle, D. R., & M. Webster.

1972 Raising children's performance expectations. Social Science Research, 1 (June): 147-158.

1974 Raising children's expectations for their own performances: a classroom application. Chapter 7 in J. Berger, T. Conner, and M. H. Fisek (eds.), Expectation States Theory: A Theoretical Research Program. Cambridge: Winthrop Publishers.

Kelman, H. C.

1961 Processes of opinion change. Public Opinion Quarterly, 25: 57-78.

Rist, R. C.

1970 Student social class and teacher expectation: the self-fulfilling prophecy in ghetto education. Harvard Educational Review, 40: 411-451.

Schachter, S.

1951 Deviation, rejection, and communication. Journal of Abnormal and Social Psychology, 46: 190-207.

Seaver, W. B.

1973 Effects of naturally induced teacher expectancies. Journal of Personality and Social Psychology, 28 (December): 333-342.

Sherif, M., B. J. White, and O. J. Harvey

1955 Status in experimentally produced groups. American Journal of Sociology, 60: 370-379.

Symonds, P. M.

1925 Notes on rating. Journal of Applied Psychology, 7: 188-195.

Whyte, W. F.

1943 Street Corner Society. Chicago: The University of Chicago Press.